

Laboratorio di Algoritmi e Strutture Dati

Docenti: M. Torelli, S. Aguzzoli

Appello del 9 Gennaio 2007

Progetto “Giacimenti”
Consegna entro il 28 Gennaio 2007

Il problema

Una compagnia mineraria ricerca nuovi giacimenti tramite sensori sonar che scandagliano il sottosuolo. Ogni rilevamento determina una mappatura di una regione rettangolare di terreno che verrà utilizzata per progettare gli scavi.

Formalmente, il suolo è l'insieme dei punti $(x, y) \in \mathbb{Z}^2$. Dato un punto $(x_0, y_0) \in \mathbb{Z}^2$ e due interi positivi r e c , una *regione rettangolare* $Rett((x_0, y_0), r, c)$ è il rettangolo il cui vertice in alto a sinistra è (x_0, y_0) e la cui base e altezza contengono rispettivamente c e r punti (i punti nel rettangolo possono essere visti come gli elementi di una matrice $r \times c$). Più precisamente:

$$Rett((x_0, y_0), r, c) = \{ (x, y) \in \mathbb{Z}^2 \mid x_0 \leq x < x_0 + c, y_0 - r < y \leq y_0 \}.$$

L'*estensione lineare* di $Rett((x_0, y_0), r, c)$ è data dal massimo fra r e c . L'*area* della regione $Rett((x_0, y_0), r, c)$ è data dal numero di punti in essa contenuti.

Un *rilevamento* effettuato da un sensore scandaglia una regione rettangolare e associa a ogni punto del rettangolo un numero intero. A ogni rilevamento è associata come *codice identificativo* una stringa finita di lunghezza arbitraria sull'alfabeto $\{A, B, \dots, Z\}$ delle lettere maiuscole. Formalmente, un rilevamento è definito dalla quintupla

$$R = \langle (x_R, y_R), r, c, M_R, \sigma_R \rangle$$

dove:

- $(x_R, y_R) \in \mathbb{Z}^2$, $r > 0$ e $c > 0$ interi; tali valori definiscono la regione rettangolare $Rett((x_R, y_R), r, c)$ interessata dal rilevamento.
- M_R è una matrice $r \times c$ di interi che definisce i valori dei punti in $Rett((x_R, y_R), r, c)$ nel modo sotto specificato.
- σ_R è il codice identificativo di R .

Rilevamenti distinti hanno codici distinti, vale a dire che se $R_1 \neq R_2$ allora $\sigma_{R_1} \neq \sigma_{R_2}$. La matrice M_R determina una mappatura del suolo $\mathcal{M}_R : \mathbb{Z}^2 \rightarrow \mathbb{Z}$ definita da:

$$\mathcal{M}_R(x, y) = \begin{cases} M_R[y_R - y, x - x_R] & \text{se } 0 \leq y_R - y < r \text{ e } 0 \leq x - x_R < c \\ 0 & \text{altrimenti} \end{cases}$$

dove $M_R[i, j]$ denota il valore della matrice M_R alla riga i ($0 \leq i < r$) e colonna j ($0 \leq j < c$). Ad esempio, se $(x_R, y_R) = (-3, 5)$ e M_R è una matrice 10×20 , allora $\mathcal{M}_R(-3, 5) = M_R[0, 0]$, $\mathcal{M}_R(-2, 5) = M_R[0, 1]$, $\mathcal{M}_R(-3, 4) = M_R[1, 0]$, mentre $\mathcal{M}_R(0, -5) = 0$ e $\mathcal{M}_R(17, 4) = 0$.

Per decidere l'area in cui operare degli scavi esplorativi, la compagnia prende in considerazione un sistema di rilevamenti e costruisce a partire da questi una mappa del sottosuolo.

Un *sistema* di rilevamenti S è un insieme finito $S = \{R_1, R_2, \dots, R_n\}$ di rilevamenti. La *mappa* \mathcal{M}_S associata al sistema S è la funzione $\mathcal{M}_S : \mathbb{Z}^2 \rightarrow \mathbb{Z}$ tale che

$$\mathcal{M}_S(x, y) = \sum_{R \in S} \mathcal{M}_R(x, y).$$

$\mathcal{M}_S(x, y)$ è il *valore* del punto (x, y) stimato dal sistema S . Data una mappa \mathcal{M}_S e una regione rettangolare T , la *disomogeneità* $d(T)$ di T è la massima differenza fra i valori di due punti appartenenti a T . Vale a dire

$$d(T) = \max\{ \text{abs}(\mathcal{M}_S(x_1, y_1) - \mathcal{M}_S(x_2, y_2)) \mid (x_1, y_1) \in T \text{ e } (x_2, y_2) \in T \}$$

dove $\text{abs}(n)$ denota il valore assoluto di n .

La scelta della regione T di scavo va fatta in modo da massimizzare l'area di T , facendo in modo che l'estensione lineare di T e la disomogeneità di T non superino i parametri prefissati.

Più precisamente, dati:

- una mappa \mathcal{M}_S ;
- un punto di *origine* degli scavi $(x_0, y_0) \in \mathbb{Z}^2$;
- i parametri $l > 0$ (estensione lineare massima) e $d \geq 0$ (disomogeneità massima);

allora la regione T *selezionata* dalla compagnia per gli scavi esplorativi sarà una regione rettangolare di area massima fra tutte le regioni della forma $\text{Rett}((x_0, y_0), r, c)$ (per opportuni interi non negativi r e c) la cui estensione lineare è $\leq l$ e la cui disomogeneità è $\leq d$.

Si osservi che per una generica scelta di \mathcal{M}_S , (x_0, y_0) , l e d possono esistere più regioni di area massima distinte fra loro.

Esempio

Si supponga di avere definito i rilevamenti

$$R_1 = \langle (2, 5), 2, 3, M_1, A \rangle \quad R_2 = \langle (3, 4), 3, 2, M_2, B \rangle \quad R_3 = \langle (4, 5), 2, 2, M_3, C \rangle$$

dove le matrici sono:

$$M_1 = \begin{bmatrix} 6 & -2 & 10 \\ 8 & 1 & 3 \end{bmatrix} \quad M_2 = \begin{bmatrix} 4 & -3 \\ 1 & 7 \\ -3 & 2 \end{bmatrix} \quad M_3 = \begin{bmatrix} -6 & -2 \\ 8 & -1 \end{bmatrix}$$

Consideriamo il sistema di rilevamento $S = \{R_1, R_2, R_3\}$. La mappa del suolo \mathcal{M}_S è descritta nella Figura 1, dove vengono evidenziati solamente i punti con valore diverso da 0.

Consideriamo il punto di origine $P_1 = (2, 5)$. Se $d = 1$, allora per ogni $l > 0$ la regione selezionata è $\text{Rett}(P_1, 1, 1)$ contenente solamente il punto P_1 . Supponiamo $d = 7$. Allora, per ogni $l > 0$ la regione selezionata è $\text{Rett}(P_1, l, 1)$. Supponiamo $d = 10$ e $l = 5$; la regione selezionata è $\text{Rett}(P_1, 3, 5)$. Se $d = 11$ e $l = 5$, la regione selezionata è $\text{Rett}(P_1, 5, 5)$. Si noti che in tutti i casi precedenti la regione selezionata è univocamente determinata.

Consideriamo il punto di origine $P_2 = (3, 5)$ e i parametri $d = 7$ e $l = 3$. In questo caso possono essere selezionate due regioni rettangolari, ossia $\text{Rett}(P_2, 1, 3)$ e $\text{Rett}(P_2, 3, 1)$. Se invece $d = 7$ e $l \geq 4$, l'unica regione selezionata è $\text{Rett}(P_2, 1, l)$.

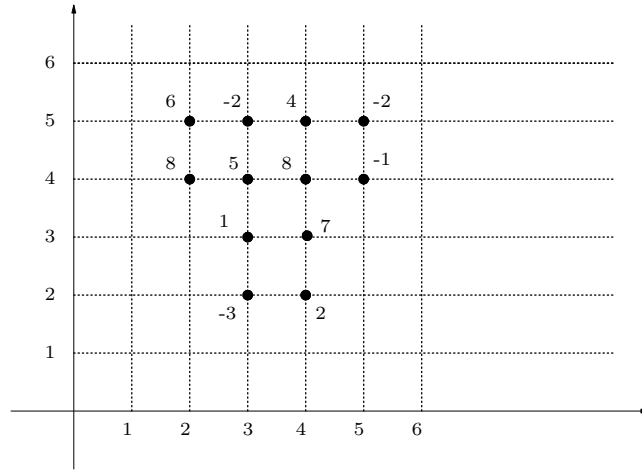


Figura 1: La mappa \mathcal{M}_S

Infine, se il punto di origine è $(-1, 2)$ e $d = 0$, allora per ogni $l > 0$ la regione selezionata è $\text{Rett}((-1, 2), l, 4)$; se il punto di origine è $(-1, 1)$ e $d = 0$, allora per ogni $l > 0$ la regione selezionata è $\text{Rett}((-1, 1), l, l)$.

Si richiede di implementare una struttura dati efficiente che permetta di eseguire le operazioni seguenti (si tenga presente che la minima porzione rettangolare di piano contenente tutti i rilevamenti può essere molto grande rispetto al numero di rilevamenti presenti nel suolo, quindi *non è sicuramente efficiente rappresentare il suolo mediante un'unica matrice*).

- **rilevamento** $(x, y, r, c, file, \sigma)$

Se non esiste già un rilevamento di codice σ , crea il rilevamento $R = ((x_R, y_R), r, c, M_R, \sigma_R)$, dove $x_R = x$, $y_R = y$, $\sigma_R = \sigma$ e i valori di M_R sono letti dal file di nome *file* secondo il formato riportato nell'apposita sezione (se il file non esiste, viene solamente stampato un opportuno messaggio di errore). Se esiste già un rilevamento di codice σ non compie alcuna operazione.

- **sistema** $(\sigma_1, \sigma_2, \dots, \sigma_n)$

Se per qualche $i = 1, \dots, n$ non esiste alcun rilevamento di codice σ_i allora non compie alcuna operazione. Altrimenti distrugge l'eventuale sistema di rilevamento esistente e crea un nuovo sistema di rilevamento $S = \{R_1, R_2, \dots, R_n\}$, dove $\sigma_{R_i} = \sigma_i$ per ogni $i = 1, \dots, n$.

- **valore** (x, y)

Se non esiste alcun sistema di rilevamento non compie alcuna operazione. Altrimenti stampa il valore del punto (x, y) stimato dal sistema di rilevamento esistente.

- **area** (x_0, y_0, l, d)

Se non esiste alcun sistema di rilevamento non compie alcuna operazione. Altrimenti stampa, secondo il formato riportato nell'apposita sezione, una regione selezionata a partire dalla mappa associata al sistema di rilevamento esistente, dall'origine (x_0, y_0) , dal parametro di estensione massima l e dal parametro di disomogeneità massima d , dove $l > 0$ e $d \geq 0$.

Si noti che le operazioni richieste sono liberamente implementabili; in particolare, non vanno necessariamente intese come prototipi di funzioni.

Specifiche di implementazione

Il programma deve leggere dallo standard input (`stdin`) una sequenza di righe (separate da `\n`), ciascuna delle quali corrisponde a una riga della prima colonna della Tabella 1, dove *file* è il nome di un file, $\sigma, \sigma_1, \sigma_2, \dots, \sigma_n$ sono stringhe finite sull'alfabeto $\{A, B, \dots, Z\}$ di lunghezza arbitraria, x e y sono interi, r, c e l sono interi positivi, $d \geq 0$. I vari elementi sulla riga sono separati da uno o più spazi. Quando una riga è letta, viene eseguita l'operazione associata; le operazioni di stampa sono effettuate sullo standard output (`stdout`), e ogni operazione deve iniziare su una nuova riga.

RIGA DI INPUT	OPERAZIONE
<code>r x y r c file σ</code>	rilevamento ($x, y, r, c, file, \sigma$)
<code>s $\sigma_1 \sigma_2 \dots \sigma_n$</code>	sistema ($\sigma_1, \sigma_2, \dots, \sigma_n$)
<code>v x y</code>	valore (x, y)
<code>a x y l d</code>	area (x, y, l, d)
<code>f</code>	Termina l'esecuzione del programma

Tabella 1: Specifiche del programma

Note

1. Non devono essere presenti vincoli sul numero e sulla dimensione dei rilevamenti (se non quelli determinati dal tipo di dato intero). Non si richiede – anzi si sconsiglia – l'uso di grafica, se non per test personali: in modo particolare, non si usi `conio.h` e neppure `clrscr()`.
2. Per semplicità si suppone che l'input sia sempre conforme alle specifiche di Tabella 1, per cui non è necessario controllare la correttezza dell'input. Per leggere l'input si usino le funzioni standard ANSI C `getchar()` e/o `scanf()`.
3. **Formato per la rappresentazione di una matrice**
Si supponga che la riga in input sia

$$r \ x \ y \ r \ c \ file \ \sigma$$

Allora il file di nome *file* deve contenere $r \cdot c$ numeri interi $a_0, a_1, \dots, a_{r \cdot c - 1}$, separati fra loro da uno o più caratteri di spaziatura (spazio, a capo, etc.), che rappresentano i valori della matrice M_R del rilevamento di dimensione $r \times c$: per ogni $0 \leq i < r$ e $0 \leq j < c$, il valore di $M_R[i, j]$ è dato dal numero $a_{c \cdot i + j}$ contenuto in *file* alla posizione $c \cdot i + j$ (nota: il primo numero del file è in posizione 0).

4. **Formato per la stampa della regione selezionata**

Si supponga che la regione selezionata dal comando

$$a \ x \ y \ l \ d$$

sia $Rett((x, y), r, c)$. Deve essere allora stampato su una nuova riga

$$x \ y \ r \ c$$

Esempio

Si supponga che le righe di input siano:

```
r -6 5 3 5 alfa.txt ALFA
r 1 4 3 2 beta.txt BETA
r -4 3 3 3 gamma.txt GAMMA
r 2 4 2 2 delta.txt DELTA
s ALFA BETA
a -6 5 100 5
a -4 4 50 18
a -4 6 10 14
a -4 4 1000 3
a -6 4 1000 7
s ALFA BETA GAMMA DELTA
v -3 3
v 2 4
a -4 4 20 11
a -6 4 20 12
a 1 4 10 5
a -4 2 10 2
a -5 4 3 5
f
```

dove il file `alfa.txt` contiene i valori

```
5 -1 3 10 -2
7 4 8 -3 5
-1 7 4 9 12
```

il file `beta.txt` contiene i valori

```
-2 5
3 1
-7 8
```

il file `gamma.txt` contiene i valori

```
-1 -5 -11
-5 4 2
-7 20 8
```

e il file `delta.txt` contiene i valori

```
-13 1
3 -5
```

L'output prodotto dal programma deve essere:

```
-6 5 2 1
-4 4 50 5
```

-4 6 3 10
-4 4 1 1
-6 4 1 3
4
-8
-4 4 2 6
-6 4 20 2
1 4 2 1
-4 2 2 1
-5 4 2 2

Presentazione del progetto

Il progetto deve essere inviato per posta elettronica all'indirizzo `aguzzoli@dsi.unimi.it` entro il 28 Gennaio 2007 (incluso).

Occorre presentare:

1. il codice sorgente (rigorosamente ANSI C, compilabile con `gcc`);
2. una sintetica relazione (formato pdf o rtf) che illustra le strutture dati utilizzate e analizza il costo delle diverse operazioni richieste dalla specifica.

I due o più file (file sorgenti C + relazione) devono essere contenuti in un unico file `.zip` il cui nome dovrà essere `cognome.zip`. La relazione e il codice devono riportare il vostro nome, cognome e matricola.

Una copia cartacea della relazione e del codice deve inoltre essere consegnata al dr. Aguzzoli entro il 29 Gennaio 2007 (lasciandola eventualmente nella sua casella postale presso il dipartimento in via Comelico).

Si ricorda infine di presentarsi alla prova orale con una copia stampata della relazione e del codice.

La discussione del progetto e l'esame orale di Algoritmi e Strutture Dati si svolgeranno **indicativamente** nei giorni 31 Gennaio, 5 Febbraio, 8 Febbraio.

Alla consegna del progetto, indicare nel *testo* della e-mail la data in cui si preferisce sostenere la prova orale; nei limiti del possibile si cercherà di tener conto di tali indicazioni (se non si hanno preferenze, non dare alcuna indicazione).

Il calendario degli esami orali sarà disponibile sulla pagina del corso qualche giorno dopo il termine di consegna del progetto.

Per ogni ulteriore chiarimento:

E-mail: `aguzzoli@dsi.unimi.it`

Ricevimento: il mercoledì, ore 15-16, stanza S204.

Avvisi

La versione aggiornata del progetto è pubblicata in `.pdf` sul sito:

`http://homes.dsi.unimi.it/~aguzzoli/algo.htm`.

Si consiglia di consultare periodicamente questo sito per eventuali correzioni e/o precisazioni relative al testo del progetto.

Si richiede allo studente di effettuare un adeguato collaudo del proprio progetto su numerosi esempi diversi per verificarne la correttezza e valutarne le prestazioni.

La realizzazione del progetto è una prova d'esame da svolgersi **individualmente**. I progetti giudicati frutto di **collaborazioni** saranno **estromessi** d'ufficio dalla valutazione.